AD ~~605496~~ G05496

MEMORANDUM
RM-3777-PR
SEPTEMBER 1964

DYNAMIC PROGRAMMIN ARNING, AND ADA ROCESSES

Richard Bellman

PREPARED FOR:

UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

MEMORANDUM

RM-3777-PR

SEPTEMBER 1964

DYNAMIC PROGRAMMING, LEARNING, AND ADAPTIVE PROCESSES

Richard Bellman

This research is sponsored by the United States Air Force under Project RAND—Contract No. AF 49(638)-700 monitored by the Directorate of Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

The RAND Corporation

1700 MAIN ST. • SANTA MONICA • CALIFORNIA • 90406

PREFACE

In this Memorandum the author indicates how the mathematical technique of dynamic programming can be used to handle a number of processes that arise in biology, engineering, economics, and psychology, and, in general, to deal with a wide class of problems that require learning and adaptation because of insufficient information about the nature of the underlying process.

SUMMARY

The intensive study in recent years of a variety of descriptive and variational processes, such as those which arise in biology, psychology, engineering, and economics, has uncovered many problems which are too complex to be solved by classical mathematical techniques. In order to describe some of the difficulties involved, the author briefly reviews the essentials of the classical approach for dealing with processes of this sort, in which there is insufficient information about the state variables. He then indicates some of the ways in which dynamic programming and adaptive control may be used to bridge the gap between classical and modern theories. Finally, the author indicates some of the problems encountered in the study of adaptive processes and suggests some directions for future research.

CONTENTS

DYNAMIC PROGRAMMING, LEARNING, AND ADAPTIVE PROCESSES

1. INTRODUCTION

The recent intensive study of biological, medical, psychological, engineering, and computer processes has uncovered large numbers of problems which escape not only solution by means of classical mathematical techniques, but even formulation.

In order to see what some of the difficulties are, it is necessary to understand the essential features of the classical approach to descriptive and variational processes. We shall briefly review the essentials of this approach and then indicate some of the ways in which dynamic programming furnishes a natural bridge between classical and modern theories.

Finally we shall indicate some of the major problems which are encountered in the study of adaptive processes and suggest some directions of research.

2. DETERMINISTIC DESCRIPTIVE PROCESSES

Let S be a physical system under examination and let us introduce a set of variables $x_1, x_2, \ldots, x_N$ describing the state of the system at any time t. The vector $x(t) = (x_1(t), \ldots, x_N(t))$ is called the state vector. To determine the behavior of the system over time, we further postulate an equation of the form

$$(2.1) \qquad \frac{dx}{dt} = g(x(s), -\infty < s \leq t),$$

where the notation indicates that the function g depends upon the entire past history of the process. In many situations, we can assume that (2.1) has the form of an ordinary differential equation

$$(2.2) \qquad \frac{dx}{dt} = g(x), \quad x(0) = c;$$

see [1] for the more general case.

The study of the properties of the system S has thus been reduced to the study of the analytic behavior of the solutions of a differential equation, a considerable reduction in difficulty.

3. STOCHASTIC DESCRIPTIVE PROCESSES

It was soon recognized that this concise description of a physical process was either not available or not applicable in a large number of significant situations. Either the functions g(x) were not known, or if precisely known, of such complicated form as to be unusable due to the high dimension of the vector x. In other cases, the initial state was not known.

To circumvent these difficulties, which at first sight appear to be major obstacles to progress, random variables were introduced, with average behavior replacing unique behavior over time.

Thus, (2.2) might be replaced by

$$(3.1) \qquad \frac{dx}{dt} = g(x(t),r(t)), \quad x(0) = c,$$

where c is a random variable and $x(t)$ is a random function of t. In some cases, as in quantum mechanics, the random variables are not explicit and the equations are of the type shown in (2.2), with the components representing probabilities or else functions from which probaoilities are generated.

4. DETERMINISTIC VARIATIONAL PROCESSES

In the study of control processes in engineering and economics, we encounter quite naturally the problem of minimizing functionals of the form

$$(4.1) \qquad J(x) = \int_0^T g(x,x',t)dt,$$

where x is subject to various initial and terminal conditions as well as to local and global constraints.

In mathematical physics, these questions arise in connection with alternative formulations of the behavior of systems.

5. DISCUSSION

In pursuing this classical route, we tacitly assume detailed knowledge of the following:

(5.1)(a) number of state variables,

(b) cause and effect,

(c) values of state variables, initially and throughout the process,

(d) probability distributions—if random variables are present,

(e) criteria—if the processes are of variational type.

How do we proceed if this information is not available?

6. LEARNING AND ADAPTIVE PROCESSES

Since we are treating new types of processes and problems, it is reasonable to expect that we will introduce some new concepts and some new analytic tools. The new concepts are those of learning and adaptation, and the new tools are dynamic programming and adaptive control. Just as the boundary between learning and adaptation is not precise, so there is considerable overlap between dynamic programming and adaptive control.

It is clear that there is little to be done about ignorance in the short run. Hence, we focus our attention upon multistage processes where information is obtained at each stage. The basic problem is that of using this information so as to improve decision making.

Fortunately, a fundamental idea from the field of engineering, namely, feedback control, provides the essential clue. A mathematical abstraction of this leads to the theory of dynamic programming [2,3,4].

With this mathematical apparatus we can handle a number of processes which arise in psychology, biology, medicine, economics, and industry—all fields where learning, adaptation, and feedback play primary roles.

The feedback to mathematics itself is in the form of new ideas and new fields in which to roam.

7. ITERATION AND TRANSFORMATIONS

Let us begin at the classical level with the concept of a transformation. Let p, a point in phase space, denote the state of a system S and let $T(p)$ denote the state a unit of time later. Then the behavior of the system over time is equivalent to the study of the iterates, $p_1, p_2, \ldots, p_n, \ldots,$ where

$$(7.1) \qquad p_1 = T(p), p_2 = T(p_1), \ldots, p_{n+1} = T(p_n).$$

8. DYNAMIC PROGRAMMING

Let us now extend this idea in the following way. Instead of keeping the transformation fixed over time, let us suppose that we have a choice of the transformation to be applied at each stage. If q denotes the choice variable, or control variable, we have

$$(8.1) \quad p_1 = T(p,q_1), p_2 = T(p_1,q_2), \ldots, p_{n+1} = T(p_n,q_n), \ldots .$$

The q_i are to be chosen so as to minimize a given criterion function

$$(8.2) \qquad R(p,p_1,\ldots,q_1,q_2,\ldots).$$

A set of q_i is called a policy, and a set which minimizes is called an optimal policy.

If we assume that R has a separable structure,

$$(8.3) \qquad R = g(p,q_1) + g(p_1,q_2) + \cdots,$$

and introduce the function

$$(8.4) \qquad f(p) = \min_{\{q\}} R,$$

then the principle of optimality [2,3,4] yields the functional equation

$$(8.5) \qquad f(p) = \min_{q_1} [g(p,q_1) + f(T(p,q_1))].$$

In the continuous case, the analogue of (8.5) yields as a by-product the Euler equation and the entire set of classical conditions of the calculus of variations [5].

9. ABSTRACTION AND EXTENSION

Since we have carefully avoided defining the phase space to which p belongs, nothing prevents us from taking p to be a point in an infinite-dimensional space or from choosing as components of p probability distributions, past histories, and so on.

We thus have a quite general formulation of multi-stage decision processes. It remains to apply this formalism to the study of learning and adaptive processes.

10. ADAPTIVE PROCESSES AND LEARNING

The fundamental tool for treating ignorance is probability theory. If we do not know the value of a parameter, we assume that it is a random variable with a given probability distribution. If we do not know the probability distribution, we take it to be a random probability distribution, an element of a family of probability distributions. If we do not know the family ... and so on. In this way, we are led quite naturally to the consideration of hierarchies of uncertainties; see the discussion in [6].

The generalized state of a system S in an adaptive process consists then not only of the usual physical state, but contains also the best current estimates of unknown quantities. These estimates may be numbers, e.g., expected values and variances, or they may be probability distributions.

At each stage of the decision process we must make a decision, a choice of q, and we must estimate the new state $T(p,q)$ on the basis of new information. Note that in many cases, part of the decision process is the determination of how much effort is to be devoted to obtaining additional information.

For analytical details, see [7], [4].

"Learning" can now be interpreted on several levels, consistent with the concept of hierarchies of uncertainty.

It is first of all the ability to estimate efficiently at each stage so that ultimately the unknown elements become known. It is secondly the ability to estimate inabilities-to-estimate on the basis of a model of simple uncertainties, and to introduce more sophisticated uncertainties, and so on.

We see then that we are led to the concept of levels of intelligence, an idea which is quite important in connection with the construction of automata.

11. APPLICATIONS

Let us note that these ideas can be applied to the construction of simulation processes, both in the business area [8] and in the field of psychiatry [9]. They afford a simple and flexible framework for the study of many multistage processes and have many immediate uses in modern control theory [4].

12. COMPLEXITY

As far as obtaining numerical answers to numerical questions is concerned, we are nowhere near a satisfactory situation. If the dimension of p is small, we have efficient routine techniques using digital computers; if the dimension is large, e.g., 10, or if p has components which are functions, these methods fail. Although a number of approximate methods exist which enable us to treat many additional classes of problems—e.g., polynomial approximation, stochastic approximation--we have not really come to grips with complexity.

In particular, we have no idea at the present of how the human mind handles situations involving huge masses of data, conflicting information, and imprecise criteria, and then makes a decision.

It seems quite clear that when we someday understand the neurophysiological basis of the human memory, or memories, and the human data-retrieval system, then we shall make progress in other areas. Furthermore, when we agree to emancipate ourselves from the restriction of universally true theorems and theories and study approximations in logical space, then we shall develop powerful approximation methods in science.

REFERENCES

1. Bellman, R., and K. L. Cooke, *Differential-Difference Equations*, Academic Press Inc., New York, 1963.

2. Bellman, R., *Dynamic Programming*, Princeton University Press, Princeton, New Jersey, 1957.

3. Bellman, R., and S. Dreyfus, *Applied Dynamic Programming*, Princeton University Press, Princeton, New Jersey, 1962.

4. Bellman, R., *Adaptive Control Processes: A Guided Tour*, Princeton University Press, Princeton, New Jersey, 1961.

5. Dreyfus, S., "Dynamic Programming and the Calculus of Variations," *Journal of Mathematical Analysis and Applications*, Vol. 1, No. 2, 1960, pp. 228-239.

6. Bellman, R., *Dynamic Programming, Intelligent Machines, and Self-organizing Systems*, The RAND Corporation, RM-3173-PR, June 1962.

7. Tou, J. T., *Optimum Design of Digital Control Systems*, Academic Press Inc., New York, 1963.

8. Bellman, R., C. Clark, C. Craft, D. Malcolm, and F. Ricciardi, "On the Construction of a Multi-person, Multistage Business Game," *Operations Research*, Vol. 5, 1957, pp. 469-503.

9. Bellman, R., M. B. Friend, and L. Kurland, *Psychiatric Interviewing and Multistage Decision Processes of Adaptive Type*, The RAND Corporation, RM-3732-NIH, August 1963.